Epidemiology for the Clinical Epidemiologist

Danielle M. Dell

2004 NEHC Workshop

Lecture Objectives

- Review epidemiological study types.
- Discuss internal and external validity, as well as bias.
- Provide an overview of choosing an appropriate test of statistical significance.
- Advise on data management and statistical packages.

Types of Epidemiological Studies

Types of Epidemiological Studies

- Case Reports & Case Series
- Ecologic studies
- Cross-sectional studies
- Observational studies
- Intervention studies

Case Reports & Case Series

- Case reports – descriptive summary of clinical observations for one patient.

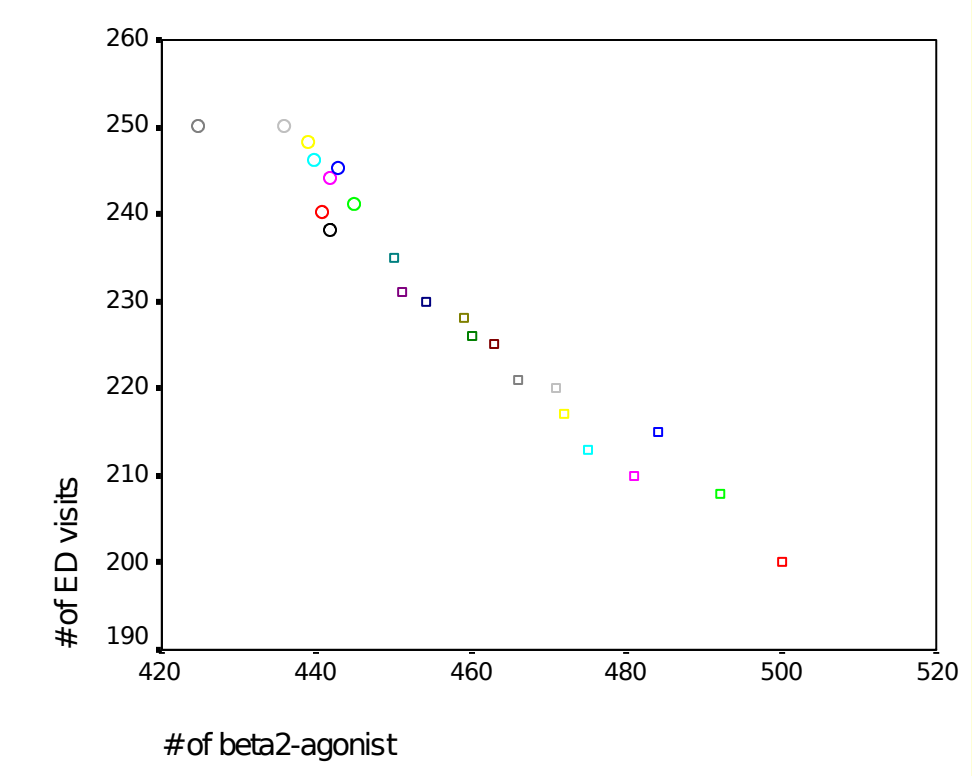
  Case reports are the most basic type of epidemiological study.

- Case series – descriptive summary of clinical observations for a group of patients with the same disease.
- These “unique” clinical observations are often the first documentation of emerging or new diseases.

Ecologic Studies

- Aggregate information for populations.
- No individual level data.
- Correlation coefficients are typically calculated.
- Cannot determine if the people who are exposed to the “risk factor” are the same people who have the “disease”.

Ecologic Study Example



- From 2002-2003, total monthly pharmacy prescriptions filled at NMCP for short-acting inhalers decreased by 15%.
- During this same time, total monthly visits to the ED for severe asthma exacerbations increased by 25%.
- Is there a relationship between the two?
- Spearman's correlation=-0.978, $p<0.001$; therefore there is a significant negative correlation between the two variables.

Cross-sectional studies

- Information about individuals is collected.
- Data on exposure and disease is collected for same point in time.
- Underlying assumption that exposure has remained constant over time.
- Relatively easy, in terms of time and cost.

Observational Studies

Cohort Studies

- Individuals are enrolled on the basis of exposure.
- Follow-up to determine outcome.
- Relative risks are calculated.

Case-control studies

- Individuals are enrolled on the basis of disease.
- Exposure is determined after disease status is known.
- Odds ratios are calculated.

Observational Studies: Advantages and Limitations

Cohort

- Good design for rare exposures.
- Temporal causal criteria satisfied.
- Can be expensive and cumbersome depending on length of follow-up.
- Loss to follow-up can introduce bias.

Case control

- Good design for rare diseases.
- Allows for the investigation of more than one risk factor.
- Recall and observation bias may be a problem.
- Odds ratio only “estimates” disease risk.

Observational Studies: Examples

Cohort

- Nurses' Health Study prospective cohort study of 120,000 U.S. nurses.
- Recently published data (Michels et al, 2003) includes 2.3 million years of person-time and demonstrated small, significant increase in risk for breast cancer among women diagnosed with type 2 diabetes.

Case Control

- MspI CYP1A1 polymorphism associated with higher risk of lung cancer among smokers in a Japanese study (Nakachi et al, 1993); cases with the metabolic faster form of CYP1A1 exhibited higher odds for lung cancer at lower pack-years than WT counterparts.

Observational Studies: The Nested Case-Control

- Case-control study conducted within a cohort.
- During follow-up the disease develops; cases and a sample of controls are selected for a case-control study.
- Particularly useful for studies that include expensive diagnostic tests.

Nested Case-Control Example

- HIV/AIDS (Polk et al, 1987)
- Nested case-control study of AIDS cases among a cohort of HIV positive men.
- Five controls matched according to date of enrollment and study center.
- Identified the following clinical factors that were predictive of AIDS: decreased number of T helper lymphocytes, increased number of T suppressor lymphocytes, low level of antibody to HIV, history of sex with someone in whom AIDS had developed.

Intervention Studies

- Most closely resemble controlled experiments.
- Investigator allocates “exposure”.
- Randomization & intention to treat analysis.
- Often considered the gold standard in epidemiological studies.
- Ethical considerations.
- Expensive and time consuming.

Intervention Study: Example

- Women's Health Initiative RCT included 17,000 post-menopausal women with intact uterus; randomized to receive estrogen+progestin or nothing.
- Trial stopped in July 2002 after women in HRT arm demonstrated significantly higher risks (26% excess) for breast cancer.

Validity, Confounding, Bias

Validity

External Validity

- Also called generalizability
- How accurately do the results observed in your study sample reflect the “risk” for populations of diseased persons outside of the study sample?
- Not due to bias or confounding.

Internal Validity

- Accuracy of the risk estimates observed within the study sample.
- Affected by bias and confounding.
- Cannot have external validity without internal validity.

Confounding

- Alteration of the true association between an exposure and disease due to a third “factor”.
- Criteria to qualify as a confounder:

    1. Must be a risk factor for the disease.
    2. Must be associated with the exposure.
    3. The exposure and disease must not affect the confounder.

Confounding, con’t.

- If the suspect factor lies within the causal pathway between exposure and disease, it is NOT a confounder.

Bias

- Introduction of systematic error into the risk estimates through study design.
- Selection bias – participants are enrolled using different criteria.
- Information – data is obtained in a different manner across study groups. (ex, observer bias, recall bias, interviewer bias)

Information Bias and Misclassification

Differential misclassification

- Exposure/disease data is systematically collected differently across study groups.
- Potential for bi-directional bias of the risk estimates.

Non-differential misclassification

- Exposure/disease data is misclassified, but not in relation to the study groups.
- Typically, risk estimates are biased towards the null.

Statistical Analysis

Continuous Data

- Summarize with measures of central tendency; mean, median, mode
- Two independent samples: t-test
- Pre and post test samples: paired t-test
- More than two samples: One-way ANOVA with pair-wise comparisons.

Categorical Data

- Summarize with frequencies and proportions.
- Chi-square tests; relative risks and odds ratios.
- If trends over time are a concern: Chi-square test for trend.
- For more than 2 disease and exposure categories:Chi-square RxC contingency table.

Parametric vs. Non-Parametric

- Parametric tests assume that the sample follows a Gaussian distribution; the prior examples were parametric tests.
- Non-parametric tests do not make this assumption.
- Non-parametric tests are not as robust; however, if the data is not normally distributed, the p-values obtained from parametric statistics will be misleading.
- Most parametric tests have a non-parametric counterpart.

Risk Estimates: Cohort Studies

- Relative Risk and 95% Confidence Intervals
- RR=$\frac{a/(a+b)}{c/(c+d)}$

		Disease		
		Yes	No	Total
Exposure	Yes	a	b	a+b
	No	c	d	c+d
	Total	a+c	b+d	

Risk Estimates: Case Control Studies

- Odds Ratio and 95% Confidence Intervals
- OR=(ad)/(bc)

  Approximates the relative risk when the outcome is rare.

Case or control?

Exposure	Case	Control	Total
Yes	a	b	a+b
No	c	d	c+d
Total	a+c	b+d	

95% Confidence Intervals and p-values

- For risk estimates, 95% confidence intervals are the preferred method of reporting because the provide the reader with the level of statistical significance, as well as the accuracy of the point estimate.
- Confidence Intervals are not necessary for summary statistics, however.
- Most statistical packages will calculate both p-values and 95% confidence intervals.

Statistical Testing

- Tests of statistical significance only determine the probability that the observed association is due to chance. (i.e., test the precision of the estimate)
- Test of statistical significance do not determine whether the results are due to bias or confounding. (i.e., does not test the accuracy of the estimate, although confounding can be controlled for.)

A Note on Over-Powering a Study

- Including large samples in a study may result in the study being over-powered to detect statistical associations.
- Comparisons that may be meaningless clinically achieve statistical significance merely because the sample size is so large.
- With the Navy’s capacity to store data, this is an issue you may all have to deal with.

Statistical Computing

Statistical Packages

- SPSS: Windows based, user friendly, advanced modules must be installed to run regression
- SAS: Windows based version is cumbersome, but programming allows for more control regarding analysis
- Excel:Crap.
- Access: Excellent data storage, poor analytic options.
- EpiInfo:New windows-based version available

Statistical Packages

- SPSS and SAS are expensive. Although, if your hospital has SPSS, you most likely have a site license which means that anyone can get the program installed. (Just keep bugging your IT person….)
- SAS licensing operates on a seat by seat basis which makes it harder to get.

What should you use?

- EpiInfo.
- Free, download-able software available from the CDC.
- Approved for use and download by the Navy.
- Newest version offers windows based analytic capabilities and allows for importation of data from Access and Excel.

A final word on statistical packages

- At least 50% of any epidemiological discipline is data management.
- You must learn at least one statistical package and you must learn it well.
- Learning these packages is a hands-on experience that cannot be acquired solely in a classroom setting.

References

- Michels KB, Solomon CG, Hu FB, et al. Type 2 diabetes and subsequent incidence of breast cancer in the Nurses' Health Study. Diabetes Care. 2003 Jun; 26(6): 1752-1758.
- Nakachi K, Imai K, Hayashi S, et al. Polymorphisms of the CYP1A1 and glutathione S-transeferase genes associated with susceptibility to lung cancer in relation to cigarette dose in a Japanese population. Cancer Res. 1993 Jul 1; 53(13): 2994-2999.
- Polk BF, Fox R, Brookmyer R, et al. Predictors of acquired immunodeficiency syndrome developing in a cohort of seropositive homosexual men. N Eng J Med. 1987 Jan 8; 316(2): 61-6.